

# SCIENTIFIC REPORTS

OPEN

## Mapping temporal-network percolation to weighted, static event graphs

Mikko Kivelä<sup>1</sup>, Jordan Cambe<sup>2,1</sup>, Jari Saramäki<sup>1</sup> & Márton Karsai<sup>2</sup>

The dynamics of diffusion-like processes on temporal networks are influenced by correlations in the times of contacts. This influence is particularly strong for processes where the spreading agent has a limited lifetime at nodes: disease spreading (recovery time), diffusion of rumors (lifetime of information), and passenger routing (maximum acceptable time between transfers). We introduce weighted event graphs as a powerful and fast framework for studying connectivity determined by time-respecting paths where the allowed waiting times between contacts have an upper limit. We study percolation on the weighted event graphs and in the underlying temporal networks, with simulated and real-world networks. We show that this type of temporal-network percolation is analogous to directed percolation, and that it can be characterized by multiple order parameters.

Contact network structure plays an important role in many dynamical processes, in particular in diffusion-like phenomena<sup>1,2</sup>. Recently, the temporal properties of networks have been shown to strongly influence spreading dynamics<sup>3–5</sup>. This is (a) because spreading processes must follow causal, time-respecting paths spanned by sequences of contacts<sup>6–10</sup> and (b) because the speed and ability of spreading processes to percolate through the contact structure are affected by temporal inhomogeneities such as the burstiness of contacts, visible as broad inter-contact time distributions<sup>11–15</sup> and correlated contact times<sup>16–18</sup>. However, this picture is still lacking detail, especially when it comes to percolation.

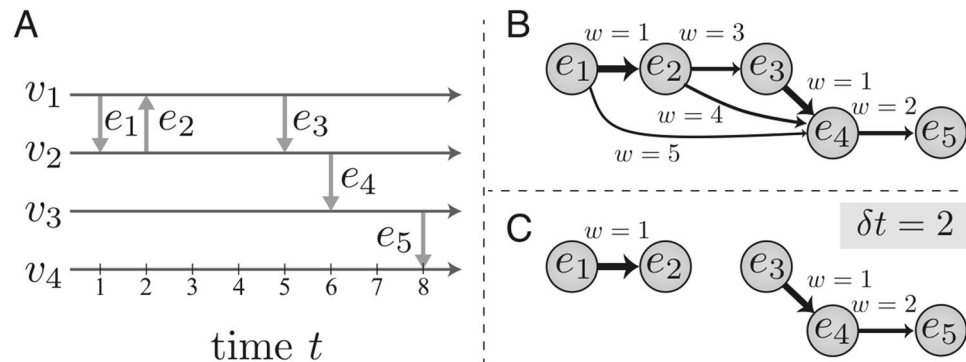
Processes with limited waiting times at nodes are particularly sensitive to broad distributions of inter-contact times; the longest inter-contact times may stop the process. Such processes include variants of epidemiological models such as Susceptible-Infectious-Recovered (SIR) and Susceptible-Infectious-Susceptible (SIS)<sup>11,19–27</sup> where nodes only remain infectious for finite periods. Other examples include social contagion<sup>28,29</sup> ad-hoc message passing by mobile agents<sup>30</sup> and passenger routing<sup>31</sup>.

In these processes, the spreading agent must be transmitted onward from a node within some time  $\delta t$  or the process stops. One can imagine a timer that starts ticking whenever the infection/message arrives at a node: this infection/message can only be transmitted onward through those of the node's contacts that happen before the timer reaches  $\delta t$ . This waiting time limit can be directly incorporated into time-respecting paths by requiring that their successive contacts are separated by no more than  $\delta t$  units of time. Because these are the only paths that the spreading process can follow, its outcome then depends on the existence of such paths. For very low values of  $\delta t$ , network-wide connectivity is unlikely and spreading processes do not percolate the network, whereas a large value  $\delta t$  may provide the pathways for infecting most of the network.

Therefore, the time limit  $\delta t$  is the control parameter of a percolation problem, where connectivity is determined by paths of contacts that follow one another within  $\delta t$ . However, discovering all paths separately for each value of  $\delta t$  is computationally expensive; a faster way would be to compute paths for a range of values, taking use of redundancy. In this article, we introduce the weighted event graph as solution to the computational problem of temporal-network percolation, and use this representation to study percolation in artificial and real networks. We show that in temporal-network percolation, there are three types of order parameters, measured in terms of component nodes, events, and lifetime and that temporal-network percolation has strong connections to directed percolation.

Weighted event graphs are static, weighted, and directed acyclic graphs (DAGs) that encapsulate the complete set of  $\delta t$ -constrained time-respecting paths for all values of  $\delta t$  simultaneously. The subset of paths corresponding to a specific value of  $\delta t$  can be quickly extracted from the weighted event graph by simply thresholding it.

<sup>1</sup>Department of Computer Science, Aalto University School of Science, FI-00076, Aalto, Finland. <sup>2</sup>Univ Lyon, ENS de Lyon, Inria, CNRS, UCB Lyon 1, LIP UMR 5668, IXXI, F-69342, Lyon, France. Correspondence and requests for materials should be addressed to M.Karsai (email: [marton.karsai@ens-lyon.fr](mailto:marton.karsai@ens-lyon.fr))



**Figure 1.** Constructing and thresholding the weighted event graph. (a) The time line of a temporal network with four nodes  $v_1 - v_4$  and five events  $e_1 - e_5$ . (b) The weighted event graph representation of the temporal network. (c) The thresholded event graph, containing only pairs of events with a maximum time difference of  $\delta t = 2$ .

Weighted event graphs can be viewed as a temporal-network extension of the line-graph representation of static networks. There is some similarity with the approach of ref.<sup>10</sup> that maps two-event sequences onto aggregated second-order networks, and with that of ref.<sup>32</sup> where an unweighted event graph is constructed from pairs of temporally closest events. Our approach builds on concepts introduced in refs.<sup>17,33</sup>.

## Results

Let us consider a temporal network  $G = (V, E, T)$  with edges defined as a set of events  $E \subset V \times V \times [0, T]$  over a time period  $T$  (e.g. see Fig. 1a). No self-edges or simultaneous events of the same node are allowed. Two events  $e = (u, v, t)$  and  $e' = (u', v', t')$  are considered adjacent so that  $e \rightarrow e'$  if they share at least one node and  $t < t'$ . This definition of adjacency is directed and preserves the arrow of time. Further, two adjacent events are considered  $\delta t$ -adjacent if their temporal distance, i.e., their time difference, is  $0 < t' - t < \delta t$ . The weighted event graph representation of  $G$  is defined as the graph  $D = (E, E_D, w)$  where the set of nodes  $E$  is the set of events in  $G$  and the edges  $e_D \in E_D$  represent the adjacency of the events  $e_D = e \rightarrow e'$  with weights defined as temporal distances  $w(e_D) = t' - t$  (see Fig. 1b). That is,  $D$  is a directed acyclic graph with links weighted with temporal distances, contains all time-respecting paths in  $G$ . For paths with a waiting time limit  $\delta t$ , we get the subgraph  $D_{\delta t}$  by thresholding  $D$  so that only links with  $w \leq \delta t$  are retained (see Fig. 1c).

The  $\delta t$ -thresholded event graph  $D_{\delta t}$  is a superposition of the time-respecting paths that a  $\delta t$ -limited spreading process can follow. Therefore, its structure tells if the process can percolate the network. A closer look at the problem reveals that here, the concept of percolation is more complex than for static networks. The components of  $D_{\delta t}$  are directed, (even if the events of  $G$  are undirected). There are only weakly connected components—there are no strongly connected components because  $D_{\delta t}$  is by definition acyclic. Each event graph node has an in-component and out-component that contain events on up- and downstream temporal paths; these components may overlap for different event graph nodes<sup>34</sup>. In the following, we will limit our analysis to weakly connected components because of their uniqueness in  $D_{\delta t}$ . For demonstration we computed the evolution of the largest in- and out-components as the function of  $\delta t$  for one empirical network as discussed in the Supplementary Materials (SM). For a spreading process to percolate, the existence of a weakly connected component is necessary but not sufficient.

The weighted event-graph representation allows us to employ computational tools developed for static networks to track percolation on temporal networks. We can sweep through the whole range of  $\delta t$  by starting from an empty graph and constructing each  $D_{\delta t}$  by adding links in the increasing order of their weight  $w(e_D)$ , and at the same time keeping track of the weakly-connected components after each addition (for further discussion see Materials and Methods). The computational cost of such process is dominated by the sorting of the edges in  $D$ . Their number scales in the worst case as  $\mathcal{O}(|E|s_{\max})$ , where  $s_{\max}$  is the maximum number of events containing the same node in the temporal network (for details on computational complexity, see Materials and Methods). One could also use conventional algorithms for temporal networks<sup>35</sup> for example to calculate the exact maximum out-component size for each value of  $\delta t$  by starting a breadth-first search for each node before each of its events and for each possible value of  $\delta t$  separately. This would, however, be extremely slow with worst-case scaling of  $\mathcal{O}(|E|^3)$ . The dramatical increase in calculation speed that weighted event graphs provide comes at a cost: the weakly connected component size only gives an upper bound for the maximum out-component (and in-component) size and the memory requirements increase as one needs to store the weighted event graph  $D_{\delta t}$ .

In percolation analysis, the relative size of the largest connected component is defined as the order parameter. Here, there are three ways of measuring the size of a component of  $D_{\delta t}$ . (1) One can count the number of event graph nodes  $S_E(E') = |E'|$  in a connected component  $E' \subseteq E$  of  $D_{\delta t}$ . This gives an upper bound for the number of events on the time-respecting paths that a spreading process can follow if it includes an event from that component. (2) One can count the number of temporal-network nodes  $S_G(E') = |\cup_{(u,v,t) \in E'} (u \cup v)|$  that are covered by the event graph component  $E'$ . This is an upper bound for the number of temporal-network nodes that any spreading process can reach via the component's time-respecting paths. Note that a temporal-network node can belong to multiple event-graph components; this can result in multiple giant components that cover most nodes but are separated in

time. (3) One can measure the *lifetime* of the event graph component  $S_{LT}(E') = (\max_{(u,v,t) \in E'} t - \min_{(u,v,t) \in E'} t)$ . This is an upper bound for the lifetime of any spreading process on the component. Note that there can be many co-existing components with long (or infinite) lifetimes; frequent and sustained contacts between a small number of nodes can already induce such components.

With these measures, we can define the order parameter as the relative size of the largest connected component,

$$\rho_*(D_{\delta t}) = \frac{1}{N_*} \max_{n_{S_*} \neq 0} S_*, \quad (1)$$

where  $n_{S_*}$  is the number of components of size  $S_*$  for the chosen definition of size  $*$   $\in \{E, G, LT\}$ , and  $N_*$  is the maximum possible value that  $S_*$  can get as a single component, i.e.,  $N_E = |E|$ ,  $N_G = |V|$ , and  $N_{LT} = T$ . In conventional percolation analysis, the average size of the other connected components is a quantity of interest that is equivalent to magnetic susceptibility. It can be introduced for the  $S_*(E')$  event graph components in  $D_{\delta t}$  as

$$\chi_*(D_{\delta t}) = \frac{1}{N_*} \sum_{S_* < \max S_*} n_{S_*} S_*^2. \quad (2)$$

One would expect this quantity to have a maximum at the critical  $\delta t_c$ , where the percolating connected component emerges in the event graph; in the thermodynamic limit this maximum would become a singularity. However, this quantity might behave differently for  $S_G(E')$  and  $S_{LT}(E')$  components due to  $\sum n_{S_*}$  not being a conserved quantity, and because of the possible multiplicity of giant components in these representations.

Note the link to *directed percolation*<sup>36</sup>, where there are two correlation lengths, temporal and spatial, characterizing correlations parallel and perpendicular to the directed lattice. In our case, the arrow of time defines the direction. However, instead of the regular lattice typical for directed percolation, our process unfolds on an irregular structure determined by the set of events that take place at each point in time. In this setting  $\rho_E$  gives the probability that a randomly selected event in  $D_{\delta t}$  belongs to a structurally percolating infinite cluster, while  $\rho_{LT}$  is the typical temporal correlation length for a given  $\delta t$ . In our case these correspond to two different order parameters, as the largest and most long-lived components might not be the same, unlike for directed percolation. Note that although we operate with the weakly connected component of  $D_{\delta t}$  for computational reasons, it is still embedded in time, and therefore it conserves the strongly anisotropic nature of the percolation process. The weakly connected component provides an upper estimate of the size or duration of the largest out-component, that is, an upper bound in terms of any of the order parameters.

**Weighted event graphs of modelled temporal networks.** To explore how  $\delta t$  controls temporal-network connectivity, we introduce a simple toy model. We define an ensemble of temporal networks  $\mathcal{G}_{p,r}(n, k, \alpha)$  where the topology is that of an Erdős-Rényi (E-R) random graph with  $n$  nodes and average degree  $k$ , and events are generated on each link by a Poisson process with  $\alpha$  events per link on average. We set the observation period  $T$  long enough so that  $\delta t \ll T$  and  $\alpha \ll T$ .

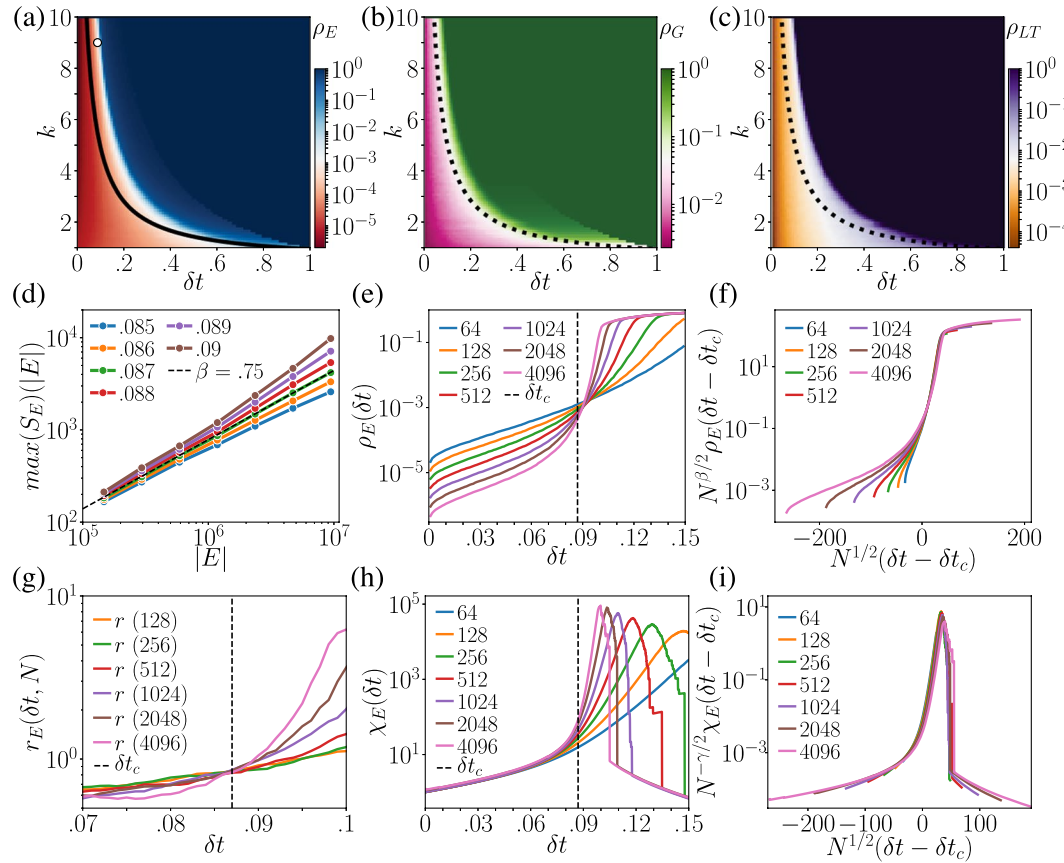
In this model, there is a transition from the disconnected to the connected phase when the independent Poissonian events become  $\delta t$ -adjacent and form a giant weakly connected component in  $D_{\delta t}$ . In terms of degree, a lower bound for this critical point can be estimated as the point where the average out-degree of the event graph becomes  $\langle k_{D_{\delta t}}^{\text{out}} \rangle = 1$ . In the underlying E-R network, each edge is adjacent to  $2(k-1) + 1$  edges (including the edge itself), and therefore the average out-degree of  $D_{\delta t}$  is  $\langle k_{D_{\delta t}}^{\text{out}} \rangle = \alpha \delta t [2(k-1) + 1]$ . The condition for the critical point can then be written as

$$k_c = \frac{(\alpha \delta t)^{-1} - 1}{2} + 1 \text{ and } \delta t_c = \frac{1}{\alpha(2k-1)}. \quad (3)$$

This theoretical line  $\delta t_c(k)$  is shown together with simulated results in Fig. 2a, with the number of events determining the relative size of the largest component,  $\rho_E$ .  $\delta t_c(k)$  separates the simulated percolating and non-percolating regimes well. Figure 2b and c show the relative largest component sizes in terms of temporal-network nodes ( $\rho_G$ ) and component lifetime ( $\rho_{LT}$ ); a percolation transition appears to take place near the theoretical line  $\delta t_c(k)$  for the number of events from Eq. (3). Note generally, the phase transition lines for events, nodes, and lifetime can be different.

Let us investigate the model's critical behavior in detail, fixing the average degree to  $k=9$ . This makes the thresholded event graph  $D_{\delta t}$  dense enough for the mean-field (MF) approach; the MF approximation works well for regular lattices above the critical dimension  $d_c=5$ . We locate the critical point with two methods. First, when the system reaches a stationary state where the order parameter becomes time-invariant beyond fluctuations, the scaling relation  $\max(S_E) \sim |D_{\delta t}|^\beta$  is expected to hold around the critical point  $\delta t_c$ , where  $|D_{\delta t}|$  is the size of the thresholded event graph in events, and  $\beta$  is the critical exponent of the order parameter. We measured this relation for several system sizes and values of  $\delta t$  and found a power-law scaling of  $S_E(|D_{\delta t}|)$  around  $\delta t_c \simeq 0.087$  with the exponent  $\beta \simeq 0.75$  (see Fig. 2d). This point is shown as a circle in Fig. 2a; it is above the analytical estimate, which provides the lower bound for the critical point. Note that for the directed-percolation universality class, the MF solution suggests  $\beta_{MF}=1$ .

The second way of determining the critical point is to calculate the ratios  $r(\delta t, N) = \rho_E(\delta t, N)/\rho_E(\delta t, N/2)$  for varying  $N$ <sup>37</sup>. These curves should cross around the critical point  $\delta t_c$  where  $r(\delta t_c, N) = 2^{-\frac{x}{\beta}}$ , and  $x$  is related to the



**Figure 2.** Phase diagrams for the random temporal network model as a function of the average network degree  $k$  and the maximum waiting time between events,  $\delta t$ . The color maps show the (ensemble-averaged) relative size  $\rho(k, \delta t)$  of the giant weakly connected components, measured as (a) the number of events in the event graph components  $S_E$ , (b) the number of temporal-network nodes that the largest event graph component covers, and (c) the lifetime of the event graph component  $S_{LT}$ . The solid line in (a) (dashed lines in (b) and (c)) is the analytic estimate of Eq. (3). The circle in the upper left corner shows the critical point for  $k=9$  determined as explained in the text. (d) Scaling of  $\max(S_E)$ , the size of the largest weakly connected component in  $D_{\delta t}$ , with the size of  $D_{\delta t}$  measured in number of event-nodes  $|D_{\delta t}| = |E|$ , for different  $\delta t$ . The dashed line assigns the critical  $\delta t_c = 0.87$ . (e) The order parameter  $\rho_E(\delta t)$  for different network sizes  $N = |V|$  with  $\delta t_c$  shown as a dashed line. (f) Same as (e) after finite-size scaling using the function defined in Eq. (4). (g) The ratios  $r(\delta t, N)$  crossing at  $\delta t_c$ . The dashed line shows the critical point determined in (d). (h) Susceptibility curves  $\chi_E(\delta t)$  for different sizes with  $\delta t_c$  shown as a dashed line. (i) Same as (h) after finite-size scaling using the function defined in Eq. (5). Computations for (a–c) are for a model network of  $|V| = 2048$  nodes evolving for  $T = 512$  time units with an event rate of  $\alpha = 1$  averaged over 10 realizations. Results for (d–i) have the same parameters but are averaged over 100 realizations and may differ in size.

finite-size scaling exponent. In Fig. 2g, they indeed cross close to  $\delta t_c \simeq 0.087$  with  $r(\delta t) \simeq 0.82$  suggesting an exponent  $x \simeq 0.2863$ , should be compared to  $\beta/2 \simeq 0.375$ .

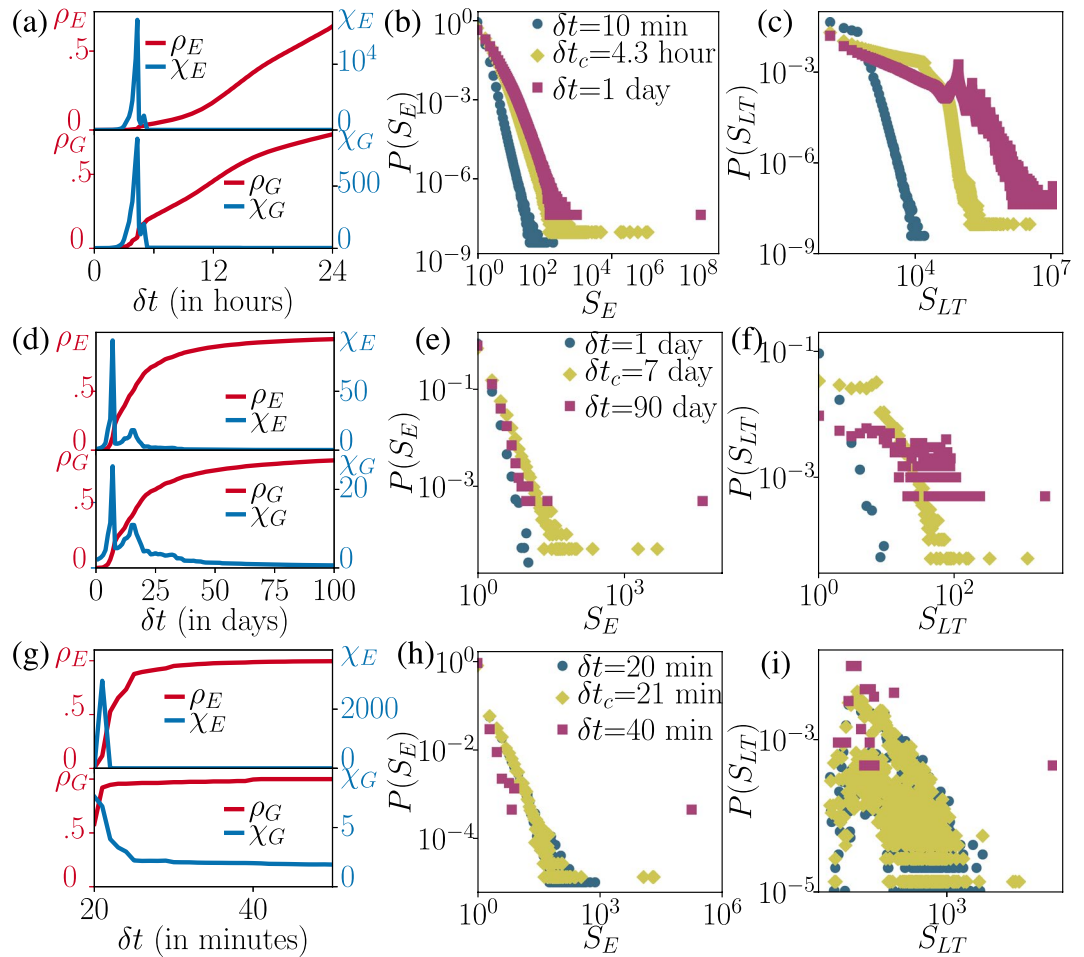
Finite-size scaling in networks is naturally related to the network volume  $N$  (number of nodes) instead of a linear size scale  $\ell$ , which usually cannot be defined. Assuming that  $N \leftrightarrow \ell^d$ , one can derive finite-size scaling functions, which are expected to hold in the conventional mean-field regime  $d > d_c$ , or for dense networks. This leads to a finite-size scaling function of the order parameter:

$$\rho_E(\delta t, N) \sim N^{-\beta/d\nu^*} \tilde{\rho}_E(N^{1/d\nu^*}(\delta t - \delta t_c)), \quad (4)$$

where  $\nu^* = 2/d$  is the finite-size scaling exponent (of linear size), which depends on the dimension  $d$ . If  $d < d_c$  it is the spatial correlation length exponent, and above the critical dimension  $d_c = 5$  it takes the value  $\nu^* = 2/d_c^{37}$ . At the same time a similar scaling function is expected to hold for susceptibility:

$$\chi_E(\delta t, N) \sim N^{\gamma/d\nu^*} \tilde{\chi}_E(N^{1/d\nu^*}(\delta t - \delta t_c)), \quad (5)$$

where  $\gamma$  is the mean cluster-size exponent. From the definition of  $\chi_E$  (in Eq. (2)) and the scaling of  $\rho(\delta t, N)$  at  $\delta t_c$  we can derive the simple exponent relation  $\gamma/(d\nu^*) = 1 - \beta/(d\nu^*)$ , where  $\nu^* = 2/d$ ,  $d = d_c = 4$  and  $\beta \simeq 0.75$ , which gives us a value  $\gamma \simeq 1.25$  (which is slightly different from the directed-percolation MF value of  $\gamma_{MF} = 1.0$ ).



**Figure 3.** Percolation transitions in empirical temporal networks of mobile communication (a–c), sexual interactions (d–f), and air transportation (g–i). Panels (a,d,g) depict the order parameter  $\rho_e(\delta t)$  (solid red lines) and susceptibility  $\chi(\delta t)$  (solid blue lines) of weakly connected components with sizes in events (upper panels) and temporal-network nodes (lower panels). Panels (b,e,h) show the size distributions of weakly connected components in the event graph measured below (blue circles), at (yellow diamonds), and above (violet squares) the critical  $\delta t_c$ . (c,f,i) are the same but depict the distribution of the lifetimes of weakly connected components in  $D$ .

To check whether the predicted finite-size scaling behaviour holds around the critical point, we took the simulated  $\rho_E(\delta t, N)$  and  $\chi_E(\delta t, N)$  measured for various  $N$  (see Fig. 2e and h respectively). Using the scaling functions in Eqs (4) and (5) with the determined exponents, we scaled the order parameter and susceptibility as a function of  $(\delta t - \delta t_c)$ . The expected scaling behaviour appears for both quantities close to the critical point (Fig. 2f and i).

**Weighted event graphs of empirical temporal networks.** We next investigated temporal percolation in real-world networks. We studied three cases: (a) a mobile call network<sup>12</sup> of  $\sim 3.2 \times 10^8$  time-stamped interactions over 120 days of  $\sim 5.2 \times 10^6$  of customers of an European operator; (b) a sexual-interaction network<sup>20</sup> from Brazil with 16,726 sex workers and clients who interacted 42,409 times over 2,231 days; and (c) an air transportation network with the time, origin, destination and duration of 180,192 flights between 279 airports in the United States<sup>38</sup> over 10 days. For details, see Materials and Methods. These networks are relevant for diffusion of information, disease, and passengers.

We measured the largest weakly connected component  $\rho_E$  (resp.  $\rho_G$ ) and susceptibility  $\chi_E$  (resp.  $\chi_G$ ) defined similarly as Eq. (2) in terms of events (resp. temporal-network nodes) covered by the event graph components. As seen in Fig. 3a for calls and in 3d for the sexual-interaction networks, the phase transitions of both types of components take place at similar times. The percolation point  $\delta t_c$  is identifiable as the peaks of susceptibility, with  $\delta t_c \sim 4$  h 20 min for the calls and  $\delta t_c \sim 7$  d for the sexual-interaction network. A spreading agent has to survive at least this long at a node to percolate the network. Interestingly, for both networks, the susceptibility shows a second peak (resp.  $\sim 5$  hours and  $\sim 16$  days), which may indicate another characteristic time-scale for the connectedness. For demonstration we computed the evolution of the largest in- and out-components for the sexual-interaction network (shown in SM), which appear with similar percolation points as the largest weakly connected component, but they evidently evolve slower in size.



For the air transportation network, we set further conditions. So far we have assumed instantaneous and bidirectional events; for flights we consider directed events with a specific duration (for details, see Materials and Methods), and require  $\delta t \geq 20$  minutes for realism (achievable transfer time). As shown in Fig. 3g this network too undergoes phase transitions, first in  $G$  at  $\approx 20$  min, and shortly thereafter in  $D_{\delta t}$  at  $\approx 21$  min. At  $\delta t = 45$  min the network becomes almost entirely connected. Note that  $\delta t = 21$  minutes is close to a minimal transfer time while 45 min can be considered as typical.

Percolation theory suggests that the structural phase transition is reflected in the component size distributions around  $\delta t_c$ . It is expected that below the critical point ( $\delta t < \delta t_c$ ), only exponentially small components are present. At the critical point  $\delta t_c$ ,  $P(S_*)$  appears with a power-law tail, and above  $\delta t_c$  the distribution is dominated by a single giant component while other components are exponentially small. This behaviour was found for weakly connected components with sizes measured as events or nodes for all three empirical systems (see Fig. 3b,e and h). However, for component lifetimes, there can be many giant components: while for small  $\delta t$ , the tail of  $P(S_{LT})$  appears as a power-law, for larger  $\delta t$  long-lived components may coexist that span the observation period but involve only a vanishing fraction of network nodes and events.

The weakly connected components discussed above only provide an upper bound for the largest reachable set of a spreading process. For a more precise measure, one to obtain the largest *out-component* which is computationally costly. This sets an important future direction: to develop an algorithmic solution for identifying the largest out-component of any node in a dense, directed, and large acyclic graph.

## Discussion

To summarize, we have introduced a new representation of temporal networks by mapping them into event graphs that are static, weighted, directed, and acyclic. Weighted event graphs recast temporal networks as static graphs that encode their topological and temporal structure without loss of information, greatly reducing the computational cost of temporal-network studies. They make it possible to use the methodology developed for directed acyclic graphs—and for time-invariant structures in general—in the analysis of temporal networks. This representation contains all time-respecting paths of a temporal network and easily yields their  $\delta t$ -constrained subset. Temporal paths are important as they determine if the diffusion of any kind of information between nodes is possible. Therefore they are pivotal for identifying the percolating structure of temporal networks as well as the outcome of any dynamical process unfolding on them. There are many examples from epidemic processes transmitted via social interactions to the routing of passengers in a transportation systems, where the timing of travel services defines the underlying temporal structure. In addition to percolation studies, weighted event graphs can be used to compute centrality scores for events, links, and nodes, and to quickly identify the complete set of  $\delta t$ -connected temporal motifs<sup>39</sup>. They can be used to study diffusion-like processes without having to compute average outcomes of stochastic simulations. Further they open new directions for studying system-level higher-order correlations in temporal networks.

## Materials and Methods

**Temporal networks with durations/delays on events.** In some temporal systems it is not enough to consider the times of events, but one needs to also consider the durations (or delays) of the events<sup>3</sup>. In this case we define a set of events as  $E \subset V \times V \times [0, T] \times [0, T]$ , where in an event  $(u, v, t, t_d) \in E$  the additional member  $t_d$  represents the duration or delay related to the event. This additional element is necessary for example in the air-traffic network studied here, where the  $t_d$  represents the flight times, and allows us to consider  $\delta t$ -adjacent time-respecting paths that can correspond to actual trips taken in the system.

The  $\delta t$ -adjacency in systems with duration or delay are defined exactly as in the simpler systems but with the allowed time difference between two events  $e = (u, v, t, t_d)$  and  $e' = (u', v', t', t'_d)$  being defined as  $0 < t' - t - t_d < \delta t$ . Note that this is not a separate definition for the  $\delta t$ -adjacency but a generalisation of the case without durations, as the  $\delta t$ -adjacency defined in the main text is returned when all events have  $t_d = 0$ .

**Definitions of  $\delta t$  adjacencies for directed networks.** In directed networks spreading, diffusion, and progress of other dynamics are constrained by the direction of the edges in addition to the arrow of time. This can be taken into account in the  $\delta t$ -adjacencies by restricting the adjacencies where  $e \rightarrow e'$  only when in the two events  $e = (u, v, t, t_d)$  and  $e' = (u', v', t', t'_d)$  have  $v = u'$ . The air-traffic network studied in the main text is considered directed in this way.

**Algorithm to construct weighted event graph representation of temporal networks.** Constructing the weighted event graph representation of a temporal network  $D = (E, E_D, w)$  can be done efficiently by noting that the edges in  $E_D$  can be listed by inspecting the sequence of events around each node  $v \in V$  separately. For some data sets the full weighted event graph  $D$  might be large, and it is convenient to construct  $D_{\delta t_{\max}}$  that can, for example, be later used to sweep through all values  $\delta t < \delta t_{\max}$ .

For each node in the temporal network  $v \in V$  one can build a time-ordered sequence of events  $\{e_1, \dots, e_k\}$  in which  $v$  participates. In the case where there are no durations one can then simply iterate over each event  $e_i$ , and for each of them search forward in the ordered event sequence until one finds an event  $e_j$  for which  $t_j - t_i > \delta t_{\max}$ . One then adds a link  $e_i \rightarrow e_j$  at the each step of this process until the event  $e_j$  that is too far from the starting event  $e_i$  is found. (Note that some  $\delta t$  adjacencies are found twice.) Creating the event sequences and sorting them can be done in  $\mathcal{O}(|E| \log |E|)$  time, and as each step of the algorithm produces a single link (with possibility of some links being visited twice) the algorithm runs in total  $\mathcal{O}(|E| \log |E| + |E_D|)$  time. Including the durations of events only

Dataset	$ E $	$ V $	$T$	Resolution
Mobile calls	324,528,907	5,193,086	120 days	1 sec
Sexual interactions	42,409	16,726	2,231 days	1 day
Air transportation	56,112	279	10 days 3 hours	1 min

**Table 1.** Summary of details about the utilised datasets.

requires a small adjustment to this algorithm, for example, a construction of sequences of events that are sorted according to the end times of the events  $t + t_d$ .

```

1: function Weighted event graph edges for a node  $\{e_1, \dots, e_k\}$ 
2:   for  $i \leftarrow 1$  to  $k$  do
3:      $j \leftarrow i + 1$ 
4:     while  $t_j - t_i \leq \delta t_{\max}$  and  $j \leq k$  do
5:       Output:  $e_i \rightarrow e_j$ 
6:        $j \leftarrow j + 1$ 

```

**Extracting component distributions from a weighted event graph.** The weighted event graph representation can turn problems related to dynamics of temporal networks to problems of graph structure, and this allows one to take advantage of computational methods developed for analysing massive graphs.

The  $\delta t$  adjacencies within a specific range  $\delta t < \delta t_{\max}$  can be calculated via “thresholding” the full network (i.e., removing all edges above the threshold level  $\delta t_{\max}$ ). In static weighed networks the weakly-connected component distributions can be calculated for all possible threshold levels very efficiently by threshold sweep, where edges are added to the network in ascending order of their weights. This typical approach in network percolation studies<sup>40</sup> can be even completed without explicitly constructing the network but by only updating a disjoint-sets forest data structure<sup>41</sup>. As the weighted event graph representation is a static weighted graph, these procedures can be used for finding the component size distribution in terms of events  $S_E$  (and quantities derived from it) for very large temporal networks and for a range of  $\delta t$  values. In practice the limit is the number of events that can be stored in memory. With some modifications, the standard algorithms can also be used to find the component size distributions in terms of the number of temporal network nodes  $S_G$  and component life times  $S_{LT}$ .

For problems related to reachability in the weighted event graph, which is a directed acyclic graph (DAG), one can simplify the full network of all  $\delta t$  connections by removing loops via transitive reduction. Removing all loops in large networks is an expensive procedure, but removing local loops, for example, around a single node in the network construction process described above is fast. Note that in the special case, when networks are undirected and no durations are present, this local procedure gives the same result as the one described recently and independently in ref.<sup>32</sup>.

There are several other ways of making computing various quantities faster that are apparent when the temporal network is represented as a weighted event graph. Any algorithms developed for DAGs, for example, to calculate reachability or transitive closures, are immediately useful in the context of temporal networks. Further, new algorithms can be easily developed with specific temporal network problems in mind. For instance, consider a case where one is interested in finding the maximum size of the out-component, which would guarantee that any process following  $\delta t$ -connectivity could not affect more than that many events in the temporal network. A naïve solution, that could be devised even without considering temporal event graphs, would be to start a breadth first search from each event in the temporal network and find the maximum. However, in the temporal event graph representation it is apparent that one can omit this search for nodes which have non-zero in-degree, and also for those belonging to weakly-connected components which are smaller than the maximum size found in previous searches.

**Detailed data description.** We utilised three datasets in this study, each recording a temporal network as a sequence of events. They were:

- Mobile call interaction network<sup>12</sup> recording 324,528,907 millions of temporal interactions over 120 days of 5,193,086 millions of customers of a single provider in an undisclosed European country.
- Sexual-interaction network<sup>20</sup> recorded in Brazil with the involvement of 16,726 sex workers and clients who interacted 42,409 times over 2,231 days.
- Air-transportation network containing the time, origin, destination and duration of 180,192 flights between 279 airports in the United States<sup>38</sup> over 10.3 days.

Details of each dataset are also summarised in Table 1. We chose these three empirical networks as they record rather different types of interactions (resp. communication, social, and transportation) and in turn are important to disseminate different types of dynamical processes (resp. the diffusion of information, epidemics, and passengers).

In data two events with no duration (i.e., zero duration) can occur in exactly the same time due to limitation in temporal resolution or due to other reason. This type of events are rare in our data sets, but they can induce loops in the network of events. In order to retain the acyclic property of these graphs we kept only a randomly selected event in case of simultaneous events of the same node.

**Data availability statement.** The sexual contact network data and public transportation data we used in Fig. 3 is openly available online<sup>20,38</sup>. The mobile call data used in Fig. 3 is shared by an undisclosed mobile operator with restrictions on availability of these data. The data was used under license for the current study, and so are not publicly available. Data are however may be available from the authors upon reasonable request and with permission of the provider.

## References

- Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical processes on complex networks*, vol. 574 (Cambridge University Press Cambridge, 2008).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).
- Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012).
- Holme, P. Modern temporal network theory: a colloquium. *Eur. Phys. J. B* **88**, 234 (2015).
- Masuda, N. & Lambiotte, R. *A Guide to Temporal Networks*, Series on Complexity Science Vol. 4 (World Scientific, 2016).
- Kempe, D., Kleinberg, J. & Kumar, A. Connectivity and inference problems for temporal networks. *Journal of Computer and System Sciences* **64**, 820 (2002).
- Moody, J. The importance of relationship timing for diffusion. *Social Forces* **81**, 25–56 (2002).
- Holme, P. Network reachability of real-world contact sequences. *Phys. Rev. E* **71**, 046119 (2005).
- Pan, R. K. & Saramäki, J. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E* **84**, 016105 (2011).
- Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nat. Commun.* **5** (2014).
- Iribarren, J. & Moro, E. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.* **103**, 038702 (2009).
- Karsai, M. *et al.* Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102 (2011).
- Kivelä, M. *et al.* Multiscale analysis of spreading in a large communication network. *J. Stat. Mech. Theor. Exp.* **2012**, P03005 (2012).
- Horváth, D. X. & Kertész, J. Spreading dynamics on networks: the role of burstiness, topology and non-stationarity. *New J. Phys.* **16**, 073037 (2014).
- Delvenne, J.-C., Lambiotte, R. & Rocha, L. E. C. Diffusion on networked systems is a question of time or structure. *Nat. Commun.* **6**, 7366 (2015).
- Pfützner, R., Scholtes, I., Garas, A., Tessone, C. J. & Schweitzer, F. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Phys. Rev. Lett.* **110**, 198701 (2013).
- Backlund, V.-P., Saramäki, J. & Pan, R. K. Effects of temporal correlations on cascades: Threshold models on temporal networks. *Phys. Rev. E* **89**, 062815 (2014).
- Aoki, T., Takaguchi, T., Kobayashi, R. & Lambiotte, R. Input-output relationship in social communications characterized by spike train analysis. *Phys. Rev. E* **94**, 042313 (2016).
- Miritello, G., Moro, E. & Lara, R. Dynamical strength of social ties in information spreading. *Phys. Rev. E* **83**, 045102 (2011).
- Rocha, L. E. C., Liljeros, F. & Holme, P. Simulated epidemics in an empirical spatiotemporal network of 50, 185 sexual contacts. *PLoS computational biology* **7**, e1001109 (2011).
- Lee, S., Rocha, L., Liljeros, F. & Holme, P. Exploiting temporal network structures of human interaction to effectively immunize populations. *PLoS One* **7**, e36439+ (2012).
- Holme, P. & Liljeros, F. Birth and death of links control disease spreading in empirical contact networks. *Sci. Rep.* **4**, 4999 (2014).
- Holme, P. & Masuda, N. The basic reproduction number as a predictor for epidemic outbreaks in temporal networks. *PLoS One* **10** (2015).
- Valdano, E., Ferreri, L., Poletto, C. & Colizza, V. Analytical computation of the epidemic threshold on temporal networks. *Phys. Rev. X* **5**, 021005 (2015).
- Génois, M., Vestergaard, C., Cattuto, C. & Barrat, A. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nat. Comm.* **6**, 8860 (2015).
- Holme, P. Temporal network structures controlling disease spreading. *Phys. Rev. E* **94**, 022305 (2016).
- Speidel, L., Klemm, K., Eguíluz, V. & Masuda, N. Temporal interactions facilitate endemicity in the susceptible-infected-susceptible epidemic model. *New J. Phys.* **18**, 073013 (2016).
- Daley, D. & Kendall, D. Epidemics and rumours. *Nature* **204**, 1118 (1964).
- Castellano, C., Fortunato, S. & Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009).
- Tripp-Barba, C., Alcaraz, C. & Igartua, M. A. Special issue on “modeling and performance evaluation of wireless ad-hoc networks. *Ad Hoc Networks* **52**, 1–2 (2016).
- Nassir, N., Hickman, M., Malekzadeh, A. & Irannezhad, E. A utility-based travel impedance measure for public transit network accessibility. *Transportation Research Part A: Policy and Practice* **88**, 26–39 (2016).
- Mellor, A. The temporal event graph. *Journal of Complex Networks*. (2017).
- Ayala, D. *Temporal percolation in the Erdős-Rényi model and the effect of burstiness*. M.Sc. thesis (Univ. Oxford, 2015).
- Nicosia, V. *et al.* Components in time-varying graphs. *Chaos: An interdisciplinary journal of nonlinear science* **22**(2), 023101 (2012).
- Wu, H. *et al.* Path problems in temporal graphs. *Proceedings of the VLDB Endowment* **7**, 721–732 (2014).
- Hinrichsen, H. Non-equilibrium critical phenomena and phase transitions into absorbing states. *Advances in Physics* **49**, 815–958 (2000).
- Karsai, M., Juhász, R. & Iglói, F. Nonequilibrium phase transitions and finite-size scaling in weighted scale-free networks. *Phys. Rev. E* **73**, 036116 (2006).
- Bureau of Transportation Statistics. [www.bts.gov](http://www.bts.gov) (2017).
- Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs in time-dependent networks. *J. Stat. Mech. Theor. Exp.* **2011**, P11005 (2011).
- Newman, M. E. J. *Networks: An Introduction*. (Oxford University Press, 2010).
- Cormen, T. H., Leiserson, C. E. & Rivest, R. V. *Introduction to algorithms*. (MIT press, 1990).

## Acknowledgements

We thank A.-L. Barabási for providing the mobile call dataset. J. S. acknowledges funding from the Academy of Finland, project no: 297195. M. Karsai acknowledges support from the Aalto Science Institute and the SoSweet ANR project (ANR-15-CE38-0011-01).

## Author Contributions

M. Kiv., J.C., J. S., and M. Kar. designed the research and participated in writing the manuscript. M. Kiv., J.C., M. Kar. implemented the algorithmic tools. M. Kar. and M. Kiv. performed the analytical and numerical analysis for the percolation study of modelled and empirical systems.



## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-29577-2>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018